Hybrid spatio-temporal structuring and browsing of an image collection acquired from a personal camera phone

Antoine Pigeau, Marc Gelgon

Hybrid spatio-temporal structuring and browsing of an image collection acquired from a personal camera phone

A. Pigeau, M. Gelgon
LINA FRE 2997 CNRS / INRIA ATLAS group, Nantes university
2, rue de la Houssinière - BP 92208, 44322 Nantes cedex 03 - France
email:{name}@lina.univ-nantes.fr

Abstract

This paper makes a proposal for automatically organizing the personal image collection that would be collected from a mobile phone equipped with a digital camera. New devices, such camera, phone camera or camcorder, enable to easily record special events of our life and involve the fast-expanding of multimedia data in everyday life. The purpose is thus to provide users efficient tools to comfortably view and retrieve information in their collection. In the present proposal, collection organization is formulated as an unsupervised classification problem, in both space and time. A criterion and an estimation procedure are proposed, based on the statistical integrated completed likelihood criterion. Then, a hybrid spatio-temporal classification is presented, obtained by fusing the temporal and geolocation-based partitions, providing an easy way to browse the collection along a single axis.

1 Introduction

The emergence of multimedia data in everyday life is fast-expanding thanks to new devices such camera, phone camera or camcorder which enable to easily record videos, sounds or images. People are used to record special events of their life and as data is gathered, it progressively builds up a valuable memory of one's life, which can be later searched for many purposes. It involves a large and growing collection of 'personal' digital media, the usage patterns of which are different from publicly available media. A new stake is thus the management of such collections in order to provide users easy access to its multimedia data.

Recently, solutions have been proposed by academics and industrials [4, 6, 8]. That take into account the specific properties of personal multimedia collections. Indeed, as discussed in [7], its collections may be distinguished from the digital library viewpoint by the content itself (nature of the scenes), the partial memory that the user has of the collection (progressive recollection is favoured by browsing if suitably organized, which is the goal of our proposal), the favorite search criteria (location, date). Specific tools are thus necessary to manage personal multimedia collection. The purpose is to provide users the possibility to comfortably retrieving a well-defined piece of information in their potentially large collection, and also functions for browsing to get an overall idea of the content of the collection. The present paper focuses on a personal image collection acquired from camera phone.

Camera phones exhibit good properties with regard to image acquisition and retrieval, as they are constantly carried around. Their permanent availability and the ability to easily share retrieved pictures (MMS) makes this context propitious for building large collections. A further quality of camera phones resides in their excellent geolocation potential, which founds our proposal. For instance, they can switch between, or fuse, measurements from an embedded GPS receiver with eg. GSM position estimation tools, such as E-OTD.

Our objective is the automated generation of a structured representation of the image collection, attempting to recover meaningful episodes and areas. This enables the user to effectively browse through time and space, yet keeping manual organisation only optional. In this paper, we solely consider time and geolocation meta-data attached to each image, while the image content itself is ignored. Further, we wish to make the scheme as unsupervised as possible, i.e. the temporal and spatial bounds of the image groups, and the number of these groups should be, as much as possible, driven by the data.

The remainder of this paper is organized as follows. Section 2 surveys existing proposals that exploit time or space for retrieval in image collections. In section 3, we propose a technique for spatio-temporal organization of one's image collection. Section 4 describes a hybrid spatio-temporal classification providing an efficient way to browse the collection. GUI aspects are also discussed. Section 5 provides experimental results. Finally, section 6 is devoted to concluding remarks.

2 Related work in time-based and geolocation-based structuring

Structuring an image collection according to the time stamp of each picture is intuitively appealing, practically quite cheap and reliable. The generative process of pictures (i.e. behaviour of users) is likely to exhibit time clusters and, furthermore, often in a hierarchical fashion. Change detection techniques, as employed in [6], possess the advantage of not setting a particular parametric model on the intra-cluster time distribution. A combination of this with clustering is proposed in [5], in which (preset size)-gap detection leads to initial groups for clustering. However, these proposals appear to come short, regarding usual issues such as number of clusters and arbitrary intra/inter-class separation thresholding. In order to cope with the variety of time scales present in the image collection, solutions such as log-scaling of inter-frame time gaps have been examined in [6]. Finally, besides direct use of time for image grouping, it was recently proposed in [3] to combine time linearly with camera settings features and image content information, within an 'image similarity' measure.

To our knowledge, there are currently few systems that seem to have considered the geolocation-based structuring closely, at least in the multimedia retrieval setting. An example is the Microsoft World Wide Media eXchange [8] system that indexes large image collections by several pieces of meta-data including location stamp. Users have the possibility to search images by location making keyword requests.

3 Spatio-temporal organization

3.1 Meta-data used and overview of the proposed approach

We formulate the recovery of the image collection spatio-temporal structure as a model-based unsupervised classification. Probabilistic model-based clustering with statistical estimation form a favourite framework for identifying meaningful groups in data [2].

In our case, the data D is assumed to be drawn from a random Gaussian mixture process with probability density

$$p(D) = \sum_{k=1}^{K} \alpha_k . \mathcal{N}(D|\mu_k, \Sigma_k), \qquad (1)$$

where the probabilities α_k are the mixing proportions and $\mathcal{N}(D|\mu, \Sigma)$ indicates a Gaussian distribution with mean μ and covariance Σ.

The main features of the proposed scheme are as follows:
1. distinct classifications are built for time and space;
2. we resort to a statistical optimality criterion, that exhibits several good properties with regard to our goal (implements Occam's razor, some degree of robustness to non-Gaussianity of clusters).
3. optimization of this criterion is conducted using a specific Expectation-Maximisation (EM) technique. Classifications are built in an incremental manner (i.e., on-line with regarding to arrival of data) using a dedicated search procedure based on merges and splits of components, and more classical EM runs.
4. the two partitions obtained are finally combined into a single, hybrid partition, occasionnally switching from one criterion to the other, depending of the presence of local structure in the data in space or time.

3.2 Optimality criterion

By taking a Bayesian hypotheses testing viewpoint, it can be shown that an effective manner of evaluating the ability of a clustering hypothesis H_K to explain the data D, taking into account the need for comparing hypotheses with various numbers of clusters, is provided by the so-called *evidence*, or marginalized likelihood:

$$P(D|H_K) = \int P(D|\Theta_K, H_K) P(\Theta_K|H_K) d\Theta_K, \qquad (2)$$

where Θ_K indicates the model parameter vector associated to hypothesis H_K. In our case, $\Theta_K = (\theta_1, \theta_2, \ldots, \theta_K)$ with $\theta_i = (\mu_i, \Sigma_i, \alpha_i)$ $1 \leq i \leq K$.
The Bayesian Information Criterion (BIC) is an effective approximation of this expression, that enables its interpretation as the data likelihood, 'automatically' penalized by model complexity. In our case, we opt for a slightly differing alternative the Integrated Completed Likelihood (ICL) criterion, presented in [1], which is well-founded, yet in practical computations comes down to penalizing the BIC criterion with the entropy of the data-to-model assignment matrix in the mixture:

$$ICL = -ML + \frac{1}{2} \cdot N(K) \cdot log(n) - \Phi(K), \qquad (3)$$

where ML is the maximized mixture loglikelihood, $N(K)$ is the number of independent parameters in the model with K components, n is the number of data elements and $\Phi(K)$ is an entropy-based criterion, defined by:

$$\Phi(K) = -\sum_{k=1}^{K} \sum_{i=1}^{n} t_{ik} \cdot log(t_{ik}), \qquad (4)$$

where t_{ik} is the posterior probability for an observation i of originating from cluster k. These t_{ik} values are supplied at convergence of the optimization phase.

3.3 Optimization of the proposed criterion

Overall, the proposed technique for optimizing the clustering criterion consists in suitable variations on the Expectation-Maximization (EM) algorithm. This local optimization framework is inherently well-suited to incremental clustering, since one may use a propagation/update mechanism as new data arrives. Further, this principle favours stability of the partitions over time, which certainly suits the wishes of the user that would appreciate some stability in the graph he navigates.

The needed ability to update the number of clusters over time (mainly increasing) suggests enabling semi-local jumps in the search space, by allowing splits & merges among clusters. We follow criteria proposed in [9], but adapt their procedure to the case of an incremental problem (Ueda et al. keep the total number of clusters constant).

Given new data, the principle of the update is to test several splits followed by several merges to update the classification. Because of the high number of split and merge possibilities, the candidates are ranked, so as to attempt splitting components with the highest entropy, as such cases suggest that the component does not fit well its associated data, or that another model also somewhat fits this data. The Mahalanobis distance is employed for tentative merging of clustering. Let us point out that, in any case, splits and merges are only carried out if they improve the ICL criterion, and their use is combined with more usual (local) improvements of the ICL, with the EM procedure. That is, the target cost function remains constant. The steps of the algorithm may be outlined as follows :

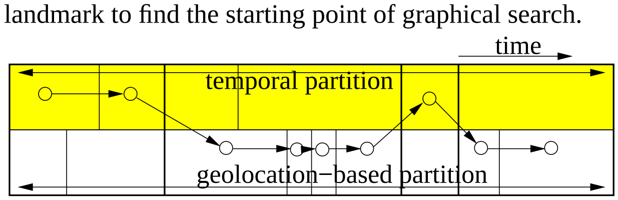
1. add new data and optimize locally using EM;
2. rank candidates for splitting according to their entropy, as computed from Φ; attempt and carry out splitting of clusters, as long as ICL is improved; run EM after each split;
3. rank candidate pairs for merging according to their Malahanobis distance; attempt and carry out merging, as long as ICL is improved; run EM after each merge.

4 Building a single, hybrid spatio-temporal partition

Let us recall that classifications are carried out independently in time and space. Some sections of the partitions obtained may not reflect in a relevant manner the structure present in the data, due to insufficiencies of the clustering criterion, poor local minima configurations, or simply because of lack of clear structure in *some part* of the data.

We propose to build a single partition out of the two obtained, trying to discard unreliable sections. As sketched in the figure below, this is formulated as the search for the best path along the time axis, possibly switching from one partition to the other at boundaries that coincide for both partitions. The principle is to obtain the hybrid partition with the strongest entropy, in the spirit of eq.(4). This is obtained at low cost, as the problem can be split into local optimizations. It is carried out as follows:
1. splitting of spatial clusters into temporally-continuous components. A spatial component can contain images taken at different temporal episodes. For each divided spatial components, we re-compute its entropy based on the t_{ik} of its associated observations.
2. selection of temporal or spatial 'blocks' with the smallest entropy. We call 'block' a time episode where limits of spatial and temporal classification coincide. The succession of the selected 'blocks' gives us the best way to browse the collection.

To browse this hybrid classification, we propose two solutions:
1. graphical interface: the user selects a date and browses the classification from the selected episode to previous or next ones. Each class is represented by several 'pertinent' images. The choice of these 'pertinent' images is not determine yet.
2. keyword request: the user has the possibility to search components by 'special events'. The problem of this approach is the annotation process which have to be carried out manually. Nevertheless, this step is facilitated since only components have to be annotate.
It has to be noted that the annotation is not an obligation since both these methods can be used simultaneously. Thus, a user can just annotate several 'special events' and use them as landmark to find the starting point of graphical search.



5 Experimental results

This experiment corresponds to a personal collection composed of 300 pictures. Figures 1 and 2 represent the classifications respectively obtained for time and location. The overall classifications being illegible for the space allowed here, we just present a zoom on the result.

The temporal classification is composed of 45 components. We find many little clusters which represent pertinent time episodes. The figure 1 shows an example of little clusters found. We can notice that badly structured data are difficult to classify as seen in component 2. The spatial classification has 25 components and relevant locations are also correctly identified. The figure 2 shows that scattered data have a tendency to be group together in large component. Several clusters may be hierarchically organized: component situated at $[17000, 16000]$ includes a little cluster near its center. It was checked that the incremental determination of clusters remains satisfactory as data arrives progressively.

The figure 3 presents the obtained hybrid spatio-temporal classification of the overall collection. The obtained partition switches between the spatial and temporal classifications: spatial $\rightarrow$ temporal $\rightarrow$ spatial $\rightarrow$ temporal. This result presents good perspective to browse the collection.

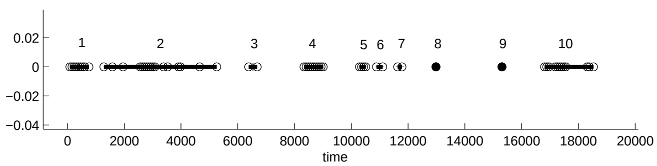


Figure 1. Zoom on the temporal classification obtained: '$+$' represents the mean of components.

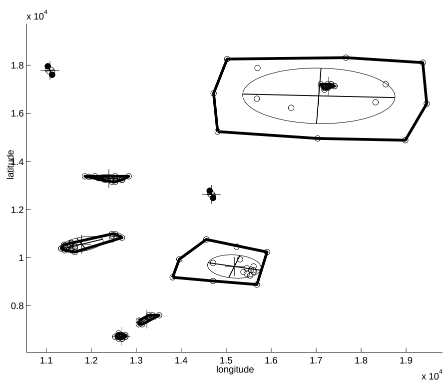


Figure 2. Zoom on the spatial classification obtained: '$+$' and ellipses represent respectively the mean and the covariance of components.

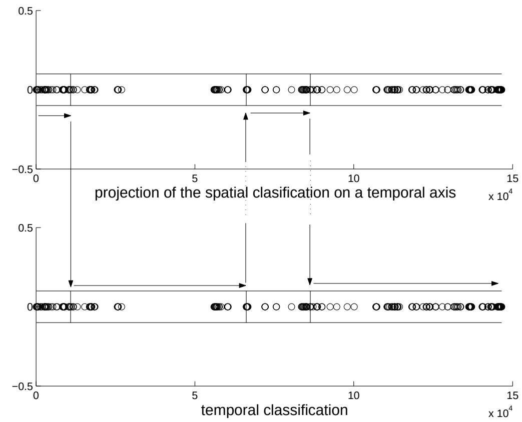


Figure 3. Hybrid classification: (top) splitting of spatial clusters into temporally-continuous components and (bottom) the temporal classification. Blocks are represented by vertical solid line and the best way to browse the data is represented by arrows.

6 Conclusion

In this paper, we focus on the problem of organizing a personal digital image collection collected from a camera phone, with a view to image browsing and retrieval. We proposed a purely data-driven structuring technique for the collection, formulated as an incremental, unsupervised classification issue. Dedicated statistical criterion and optimization procedures are designed to this purposed, that avoid delicate parameter tuning. Finally, a hybrid spatio-temporal classification is proposed to easily browse the collection. We are currently extending this scheme to structure the collection at multiple scales.

References

[1] C. Biernacki, G. Celeux, and G. Govaert. Assessing a mixture model for clustering with the integrated classification likelihood. In *IEEE Transaction on pattern analysis and machine intelligence*, volume 22, pages 719–725, Jul. 2000.

[2] C. Fraley and A. E. Raftery. How many clusters? Which clustering method? Answers via model-based cluster analysis. *The Computer Journal*, 41(8):578–588, 1998.

[3] U. Gargi, Y. Deng, and D. R. Tretter. Managing and searching personal photo collections. Technical Report HPL-2002-67, HP Laboratories, Palo Alto, Mar. 2002.

[4] J. Gemmell, R. Lueder, and G. Bell. The mylifebits lifetime store. In *Proceedings of the 2003 ACM SIGMM workshop on Experiential telepresence*, pages 80–83, Nov. 2003.

[5] A. Graham, H. Garcia-Molina, A. Paepcke, and T. Winograd. Time as essence for photo browsing through personal digital libraries. In *ACM Joint Conference on Digital Libraries JCDL*, pages 326–335, Jun. 2002.

[6] J. C. Platt and B. A. F. M. Czerwinski. PhotoTOC: Automatic clustering for browsing personal photographs. Technical Report MSR-TR-2002-17, Microsoft Research, Feb. 2002.

[7] K. Rodden. How do people manage their digital photographs? In *ACM Conference on Human Factors in Computing Systems*, pages 409 – 416, Fort Lauderdale, Apr. 2003.

[8] K. Toyama, R. Logan, A. Roseway, and P. Anandan. Geographic location tags on digital images. In *Proceedings of the eleventh ACM international conference on Multimedia*, pages 156–166, Berkeley, CA, USA, Nov. 2003.

[9] N. Ueda, R. Nakano, Z. Gharhamani, and G. Hinton. SMEM algorithm for mixture models. *Neural computation*, 12(9):2109–2128, 2000.